

## ORIGINAL ARTICLE

# Mapping the regulatory sequences controlling 93 breast cancer-associated miRNA genes leads to the identification of two functional promoters of the *Hsa-mir-200b* cluster, methylation of which is associated with metastasis or hormone receptor status in advanced breast cancer

EJH Wee<sup>1,5</sup>, K Peters<sup>1,5</sup>, SS Nair<sup>3</sup>, T Hulf<sup>3</sup>, S Stein<sup>2</sup>, S Wagner<sup>2</sup>, P Bailey<sup>1</sup>, SY Lee<sup>1</sup>, WJ Qu<sup>3</sup>, B Brewster<sup>1</sup>, JD French<sup>1</sup>, A Dobrovic<sup>4</sup>, GD Francis<sup>2</sup>, SJ Clark<sup>3</sup> and MA Brown<sup>1</sup>

<sup>1</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, Australia; <sup>2</sup>Pathology Queensland, Princess Alexandra Hospital, Woolloongabba, Queensland, Australia; <sup>3</sup>Epigenetics Group, Cancer Program, Garvan Institute of Medical Research, Sydney, New South Wales, Australia and <sup>4</sup>Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia

MicroRNAs (miRNAs) are small non-coding RNAs of ~20 nt in length that are capable of modulating gene expression post-transcriptionally. Although miRNAs have been implicated in cancer, including breast cancer, the regulation of miRNA transcription and the role of defects in this process in cancer is not well understood. In this study we have mapped the promoters of 93 breast cancer-associated miRNAs, and then looked for associations between DNA methylation of 15 of these promoters and miRNA expression in breast cancer cells. The miRNA promoters with clearest association between DNA methylation and expression included a previously described and a novel promoter of the *Hsa-mir-200b* cluster. The novel promoter of the *Hsa-mir-200b* cluster, denoted P2, is located ~2 kb upstream of the 5' stemloop and maps within a CpG island. P2 has comparable promoter activity to the previously reported promoter (P1), and is able to drive the expression of miR-200b in its endogenous genomic context. DNA methylation of both P1 and P2 was inversely associated with miR-200b expression in eight out of nine breast cancer cell lines, and *in vitro* methylation of both promoters repressed their activity in reporter assays. In clinical samples, P1 and P2 were differentially methylated with methylation inversely associated with miR-200b expression. P1 was hypermethylated in metastatic lymph nodes compared with matched primary breast tumours whereas P2 hypermethylation was associated with loss of either oestrogen receptor or progesterone receptor. Hypomethylation of P2 was associated with gain of HER2 and androgen receptor expression. These data suggest an association between miR-200b regulation and breast cancer subtype and a potential use of DNA methylation of miRNA promoters as a component of a suite of breast cancer biomarkers.

*Oncogene* (2012) 31, 4182–4195; doi:10.1038/onc.2011.584; published online 9 January 2012

**Keywords:** MicroRNA; miR-200b; promoter; DNA methylation; breast cancer; prognosis

## Introduction

Breast cancer is a heterogeneous disease that can be classified on the basis of a number of characteristics including tumour size, histological subtype and grade, oestrogen (ER) and progesterone (PR) receptor and HER2 expression, axillary lymph node (LN) status and expression profile (Sorlie *et al.*, 2001). Some of these features have been associated with disease characteristics and can therefore be used to inform patient management. For example, patients with tumours that test positive for ER and HER2 can be treated with tamoxifen and herceptin, respectively, and have a significantly better prognosis than those that test negative for these markers. However, the heterogeneity that exists even within breast cancer subgroups defined by multiple markers means that for the vast majority of breast cancer cases, predicting outcome remains a challenge, and thus additional informative biomarkers are urgently needed.

Breast cancer results from abnormalities in the quality or quantity of certain gene products, including coding and non-coding genes. MicroRNAs (miRNAs) are small non-coding RNAs ~20 nt in length that are capable of modulating gene expression post-transcriptionally (Cullen, 2004; Boyd, 2008; Bartel, 2009). MiRNAs can exhibit either tumour suppressor or oncogenic roles by modulating key cellular processes in cell-cycle progression, apoptosis and invasion (Bartels and Tsonalis, 2009; Mirnezami *et al.*, 2009; Visone and Croce, 2009). In several studies, differential miRNA expression has been shown to distinguish normal and breast tumour tissue, breast cancer subtypes, ER, PR and HER2 status, and to predict lymph node status and invasive-

Correspondence: Dr MA Brown, School of Chemistry and Molecular Biosciences, The University of Queensland, Coopers Road, St Lucia, Queensland 4072, Australia.

E-mail: melissa.brown@uq.edu.au

<sup>5</sup>These authors contributed equally to this work.

Received 13 February 2011; revised 22 April 2011; accepted 28 April 2011; published online 9 January 2012

ness (Iorio *et al.*, 2005; Mattie *et al.*, 2006; Foekens *et al.*, 2008; Yan *et al.*, 2008; Lowery *et al.*, 2009). Together, these studies suggest a potential diagnostic and prognostic use of miRNAs as biomarkers in breast cancer.

Quantitative defects in miRNAs arise through several mechanisms, including aberrant DNA methylation. Human DNA methylation usually occurs at the number 5 carbon of cytosine of a CpG dinucleotide motif. High densities of CpGs, termed CpG islands (CGIs) are usually associated with promoter elements and methylation of which usually leads to gene repression. Aberrant DNA methylation of miRNA genes has been associated with several cancers (Lujambio *et al.*, 2007; Lehmann *et al.*, 2008; Lodygin *et al.*, 2008), suggesting a possible use of miRNA DNA methylation as a prognostic tool. For example, miR-9-1 and miR-34a are hypermethylated in breast cancer (Lehmann *et al.*, 2008; Lodygin *et al.*, 2008). In addition, the miR-200b cluster (miR-200b, -200a and -429), has a CGI associated promoter ~4 kb upstream of the sequence encoding the mature miRNA (Bracken *et al.*, 2008), and aberrant DNA methylation of this sequence is associated with loss of miR-200b expression in colon (Han *et al.*, 2007), bladder (Wiklund *et al.*, 2011) and pancreatic (Li *et al.*, 2010) cancers. The contribution of miR-200b cluster gene methylation to breast cancer has not yet been reported.

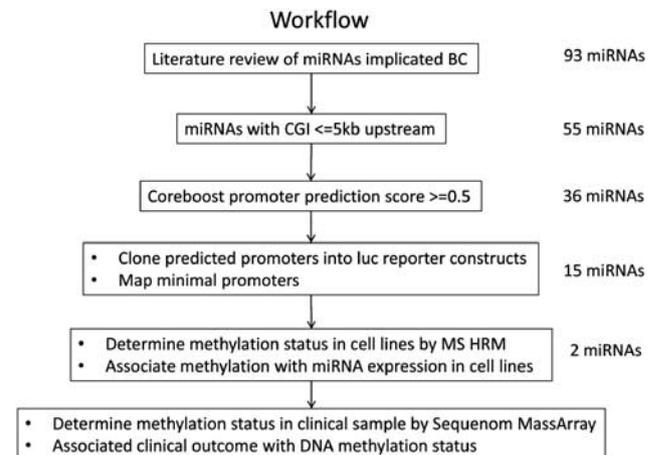
Although much is known about the biogenesis and function of miRNAs, relatively little is known about the transcriptional regulation of miRNA genes. To date, a limited number of miRNA promoters have been experimentally characterized and only recently have several miRNA promoter prediction algorithms emerged (Zhou *et al.*, 2007; Fujita and Iba 2008; Linhart *et al.*, 2008; Marson *et al.*, 2008; Oszolak *et al.*, 2008; Wang *et al.*, 2009). These studies show that miRNA promoters lie anywhere from a few bases upstream of the stemloop to tens of kilobases upstream (Linhart *et al.*, 2008; Marson *et al.*, 2008; Oszolak *et al.*, 2008; Wang *et al.*, 2009). Furthermore, several miRNAs have multiple promoters (Oszolak *et al.*, 2008; Wang *et al.*, 2009; Monteys *et al.*, 2010).

In this study, 93 miRNAs previously associated with breast cancer, were prioritized for experimental analysis using bioinformatics to look for CGI-associated promoters. The CGI-associated promoters of 15 miRNAs were mapped and methylation determined in a panel of nine breast cancer cell lines. A novel promoter for the miR-200b cluster and its role in regulating miR-200b expression was investigated. The relationship between methylation of this promoter and the previously described miR-200b cluster promoter with miRNA-200b expression and clinical characteristics in breast cancer are described.

## Results

### *Fifty-five miRNAs previously implicated in breast cancer are located within 5 kb of a predicted CGI*

To identify candidate promoters for which methylation could be associated with breast cancer development, a



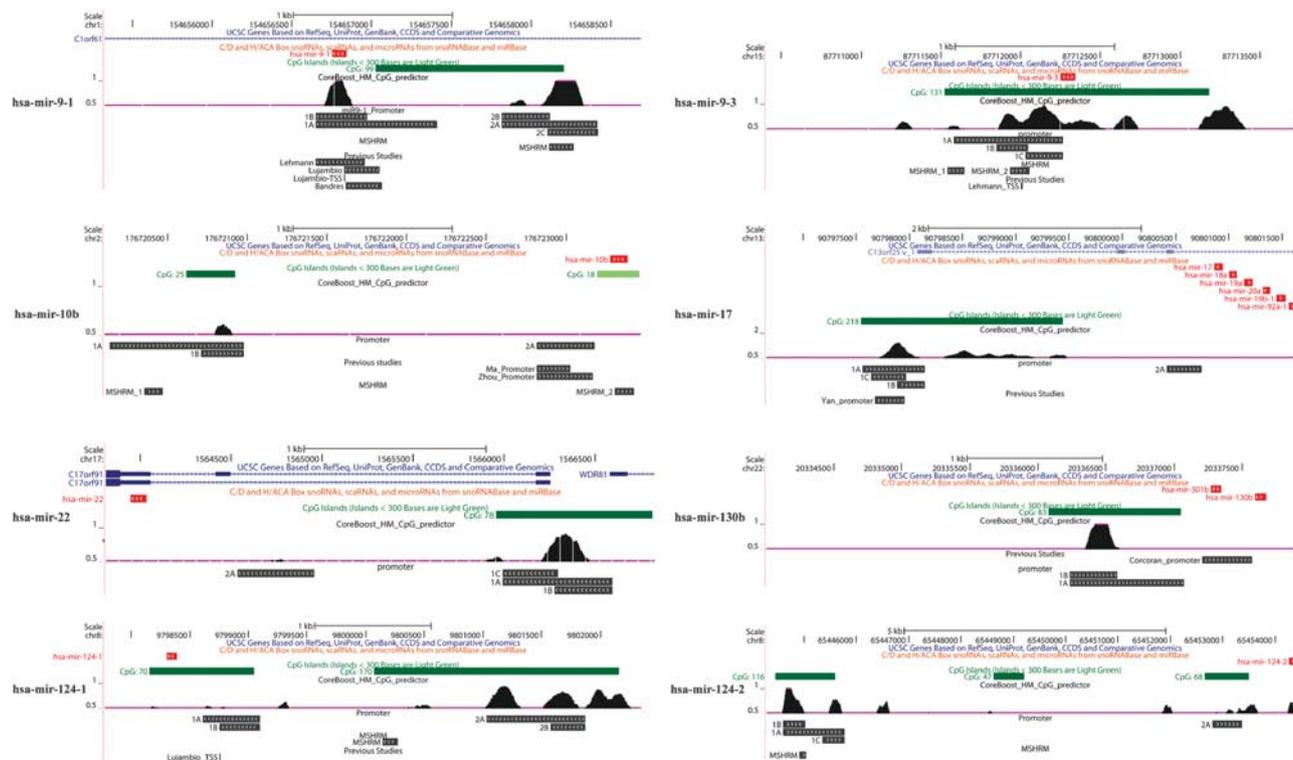
**Figure 1** Overview of the miRNA selection approach performed in this study. A literature review of miRNAs implicated in breast cancer was followed by *in silico* and molecular studies in human breast cancer cell lines. Analysis of methylation in clinical breast cancer specimens was performed on miR-200b.

list of 93 miRNAs implicated in breast cancer was collated from the literature (Supplementary Table 1). Using CpGPlot and CGI searcher, 55 (59%) of these miRNAs had a predicted CGI within 5 kb upstream of the region encoding their 5' stemloop. The Core Boost\_HM promoter prediction algorithm was used to predict regulatory elements controlling the transcription of the 55 miRNA-associated CGIs. This algorithm uses PolII chip binding, histone modifications and DNA motifs associated with promoters to predict putative transcription start sites (TSS) (Wang *et al.*, 2009). Putative promoters were defined by a CpG promoter prediction cutoff score of 0.5, representing a 90% likelihood of TSS within 500 bp of the predicted region. Figure 1 summarizes the process.

### *Experimental validation of predicted novel promoters of 15 miRNAs*

To determine if these predicted CpG promoter sequences had experimentally detectable promoter activity, a 600–1000 bp of genomic sequences around the predicted site was cloned upstream of a luciferase gene and assayed for reporter activities in either MCF7 or MDA-MBD-231 breast cancer cell lines. When a miRNA had more than one predicted promoter, a fragment encompassing each prediction was cloned. Their genomic locations are detailed in Figures 2 and 3 and Supplementary Table 1. The previously described promoters of the miR-17 cluster (Yan *et al.*, 2009) and the miR-200b cluster (Bracken *et al.*, 2008) were included as positive controls whereas a non-CoreBoost\_HM predicted fragment in miR-17 cluster was used to control for background promoter activity.

Twenty-two novel promoters from 15 miRNAs exhibited at least fivefold activity compared with the promoter-less pGL3-basic control in at least one cell line (Figures 4–6a). As expected, the previously described



**Figure 2** UCSC screenshots of miRNA candidates and their associated genomic features. Bars representing miRNAs are shown in red, CGIs in green, promoter and methylation-sensitive high-resolution melt analysis fragments in black. Annotated genes are marked in blue. Orientation of genes and fragments are indicated by directional arrows. CoreBoost\_HM promoter predictions are shown as black peaks.

promoters of miR-17 and miR-200b clusters had strong promoter activity whereas the non-CoreBoost\_HM predicted fragment had no detectable promoter function (Figures 3d and 4a). The 15 miRNAs with experimentally validated promoters are miR-9-1, miR-9-3, miR-10b, miR-22, miR-124-1, miR-124-2, miR-124-3, the miR-130b cluster, miR-193b, miR-200b cluster, miR-210, miR-320a, miR-335, miR-373 and miR-663. Three promoters were mapped for the miR-124-3 loci; two promoters were mapped for miR-9-1, 22, 124-1, 124-2, 193b and 200b; and one promoter was mapped for miR-9-3, 10b, 130b, 210, 320a, 335, 373 and 663. To map the minimal promoters regions and to facilitate methylation analysis, promoter fragments of the 15 miRNAs were fine mapped to ~300 bp (Figures 2, 3 and 6b).

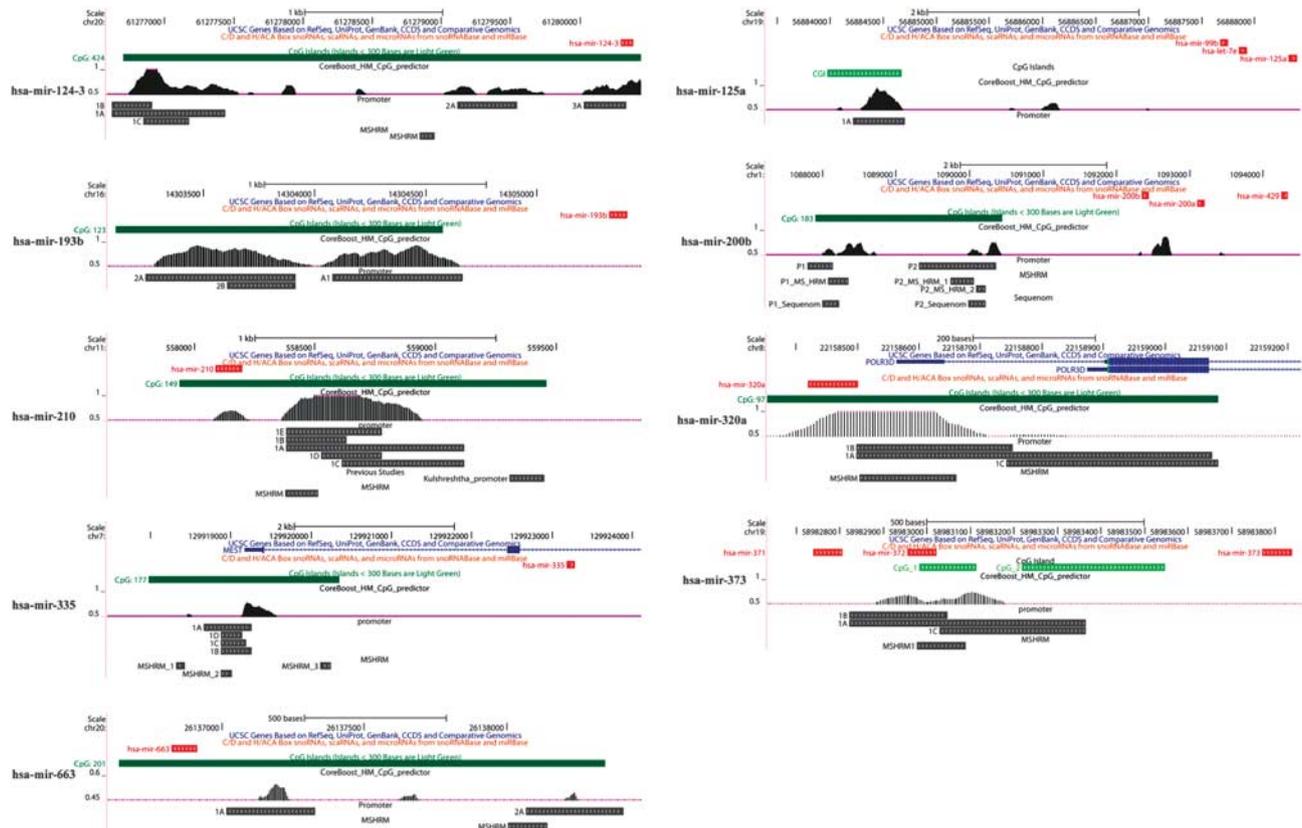
*The miR-200b cluster P2 promoter is sufficient to drive expression of miR-200b*

To determine whether the P2 promoter could drive the expression of miR200b in its endogenous genomic context, low miR-200b expressing MDA-MB-231 cells (Gregory *et al.*, 2008), were transfected with a miR-200b minigene spanning the P2, but not the P1, promoter and the sequence corresponding to the mature miR-200. This minigene was generated by replacing the luciferase coding sequence of pGL3-basic with the miR-200b genomic sequence (Figure 6c). The introduction of the miR-200b minigene resulted in an eightfold increase

in mature miR-200b expression over the pGL3-basic control (Figure 6c). Deletion of the minimal promoter in the minigene reduced miR-200b expression by 50% (Figure 6c). Collectively, these results indicate that the P2 promoter can regulate miR-200b, and very possibly miR-200a and 429, as a polycistronic primary transcript (Bracken *et al.*, 2008).

*The miR-200b cluster P1 and P2 promoters are independent*

To address the hypothesis that P1 and P2 promoters function synergistically to enhance expression of the miR-200b cluster, a 2.5-kb fragment encompassing both promoters was cloned upstream of the luciferase gene and assayed for reporter activity in MDA-MB-231 cells, in which only P2 was observed to be functional, and MCF7 cells, in which functional activity was observed for both promoters. As predicted, the P1 + P2 fragment produced similar reporter activity to the P2 fragment alone in MDA-MB-231 cells ( $P=0.34$ ) (Figure 6a). In contrast, in MCF7 cells, the reporter activity of P1 + P2 was not significantly greater than the activity of P1 ( $P=0.4$ ), but was significantly stronger than P2 ( $P<0.05$ ) (Figure 6a). Also, the activity of P2 alone and P1 + P2 were also significantly higher ( $P<0.05$ ) in MCF7 than in MDA-MB-231 cells (Figure 6a). Taken together, this data suggests that the P1 and P2 promoters function independently.



**Figure 3** UCSC screenshots of miRNA candidates and their associated genomic features. Bars representing miRNAs are shown in red, CGIs in green, promoter and methylation-sensitive high-resolution melt analysis fragments in black. Annotated genes are marked in blue. Orientation of genes and fragments are indicated by directional arrows. CoreBoost\_HM promoter predictions are shown as black peaks.

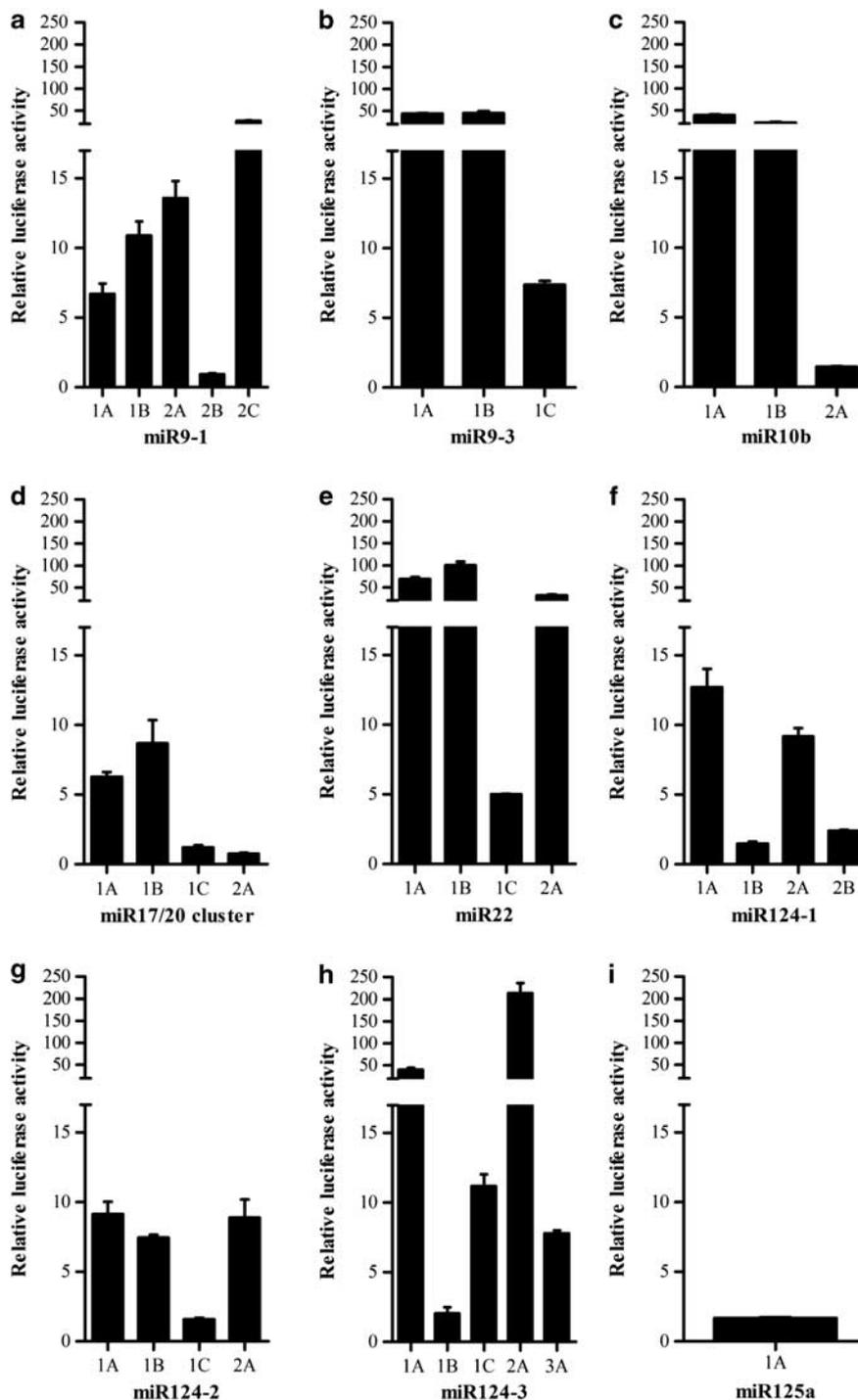
*The miR-200b cluster has multiple TSS*

To complement the promoter mapping experiments, attempts were made to map the TSS of P2. Classical 5' RACE PCR was employed to determine the TSS of P2 using RNA from MDA-MB-231 cells transfected with the P2 construct to enrich for P2 derived transcripts. However, repeated attempts with the classical 5' RACE protocol were unsuccessful, consistently producing non-specific smears (data not shown). Successful amplification of template controls indicated that the cDNA synthesis had worked and that this result was more likely to reflect heterogeneity in miR-200b cluster transcripts. An alternative 'PCR walk' approach to mapping the TSS was performed using a single transcript-specific reverse primer and various forward primers toward the 5' end of the cDNA transcript. The longest transcript extended from -3032 bp to -2447 bp upstream of the 5' stemloop as indicated by loss of PCR amplification (Figure 6d). This observation was in agreement with the both the minigene and the luciferase reporter assays. To further test the hypothesis that miR-200b has multiple TSS, publicly available breast cancer specific RNA-Sequence and RNA PolII-chip data were analysed at the miR-200b loci (Figure 6e). Multiple RNA-sequence peaks were observed along the CGI for T47D and MCF7 cells indicating expression from the CGI. Furthermore, multiple RNA PolII binding signals

in MCF7 cells were detected along the associated CGI suggesting multiple TSS. A strong RNA PolII signal overlapping P1 also suggested preferential transcription from P1 in MCF7 (Figure 6e).

*Methylation of miR-200b cluster and miR-335 promoters is associated with reduced miRNA expression*

DNA methylation of the minimal promoters of the 15 miRNA was assessed by methylation-sensitive high-resolution melt analysis (Wojdacz and Dobrovic, 2007) in a panel of nine breast cancer cell lines. The proximal miR-9-1 promoter was not included as methylation of this promoter had been previously described (Lehmann *et al.*, 2008). The miR-17 cluster promoter was also excluded because the high density of CG dinucleotides made it unsuitable for methylation-sensitive high-resolution melt analysis. Promoter methylation was then compared with miRNA expression in the same cell lines. Mir-200b cluster and miR-335 promoter methylation were inversely associated with miRNA expression. For miR-200b cluster, eight out of the nine cell lines displayed an inverse association (Figure 7a, Supplementary Figure 2). Although only MCF7 highly expressed miR-335, MCF7 also had the lowest methylation compared with the remaining eight, which were fully methylated and had minimal miR-335 expression

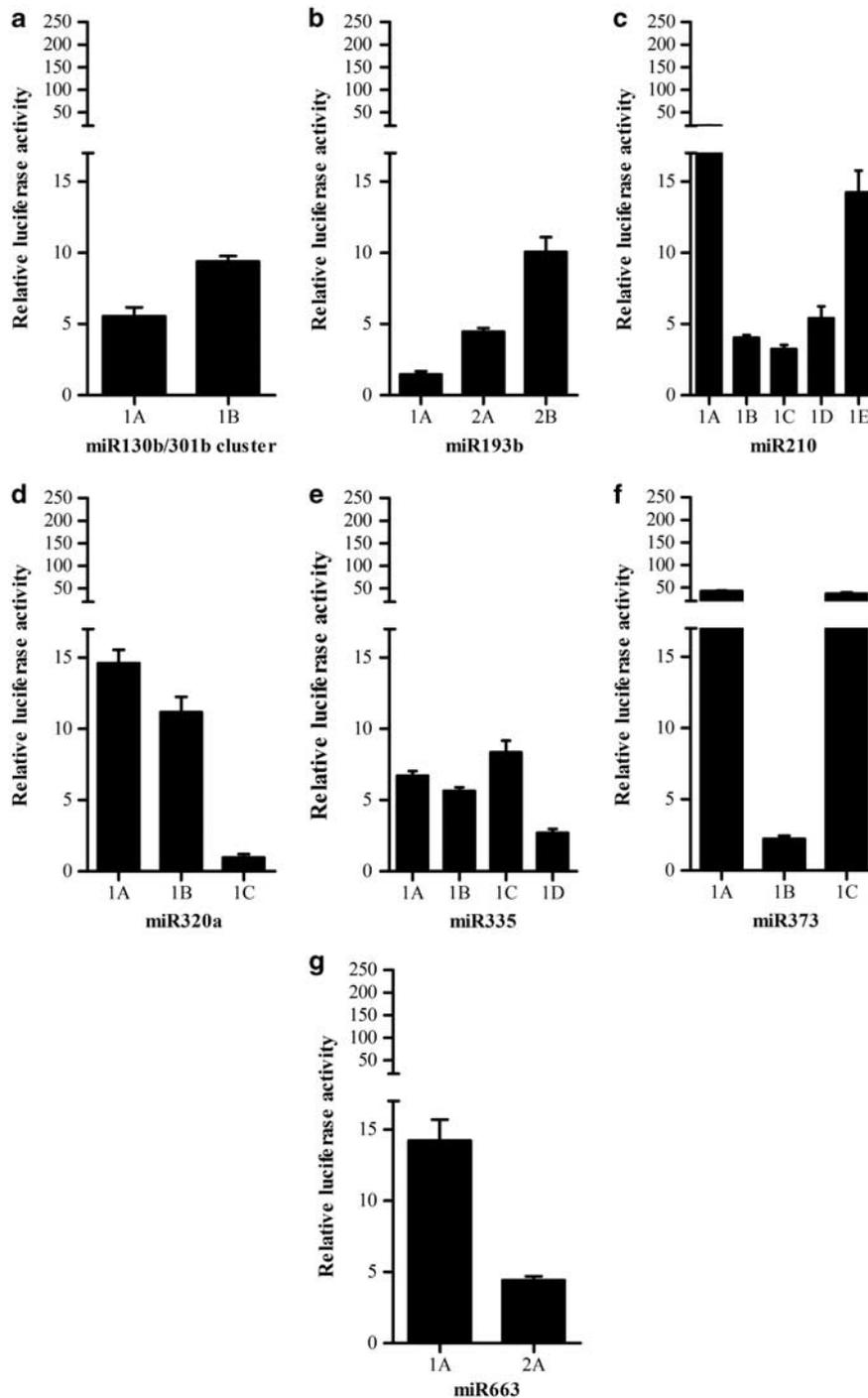


**Figure 4** (a–i) Promoter activities of miRNA candidates. miRNA promoter activity in cells expressed in RLU  $\pm$  the s.e.m. Data were generated from three independent experiments. Promoter fragments are labelled 1, 2 or 3 and A, B or C indicates the sub-fragment of that respective promoter.

(Supplementary Figure 1A). However, since the inverse association was stronger in miR-200b, P1 and P2 promoters represented better candidates for further analysis. In contrast, the miR-210 and miR-320a promoters were unmethylated in all nine cell lines although DNA methylation was not associated with miRNA expression for miR-9, miR-10b, miR-124, miR-373 and miR-663 (Supplementary Figure 1).

#### *The minimal miR-200b cluster promoters are regulated by DNA methylation*

The novel P2 promoter had comparable activity to P1 in MCF7 cells, but unlike P1, was functional in both cell lines tested (Figure 6a). The minimal P2 promoter maps to  $-2228/-1993$  bp upstream of the miR-200b 5' stemloop (Figure 6b). To confirm that DNA methylation directly repressed promoter activity, P1 and P2 were



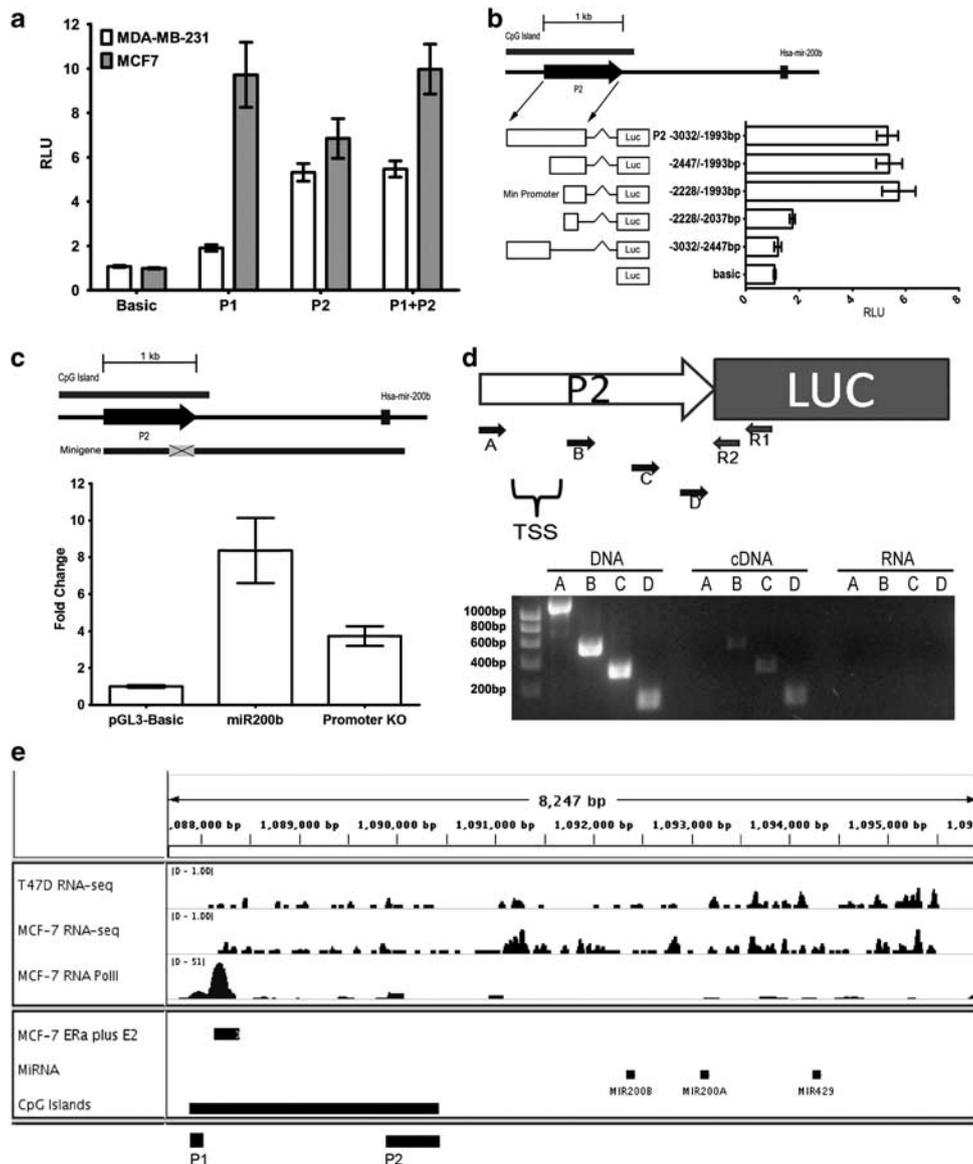
**Figure 5** (a–g) Promoter activities of miRNA candidates. miRNA promoter activity in cells expressed in RLU  $\pm$  the s.e.m. Data were generated from three independent experiments. Promoter fragments are labelled 1, 2 or 3 and A, B or C indicates the sub-fragment of that respective promoter.

cloned into a CpG-free reporter construct (Klug and Rehli, 2006) and *in vitro* methylated by *SssI* DNA methylase. Methylated P1 and P2 constructs displayed a significant reduction in promoter activity, compared with their mock methylated constructs when transfected into T47D cells, in which both promoters are endogenously unmethylated and functional (Figure 7b). This

suggests that DNA methylation represses miR200b cluster promoter activity.

*miR-200b P1 and P2 promoters are differentially methylated in primary breast tumours*

To study DNA methylation of the miR-200b promoters, Sequenom MassArray was performed on Grade 3 FFPE

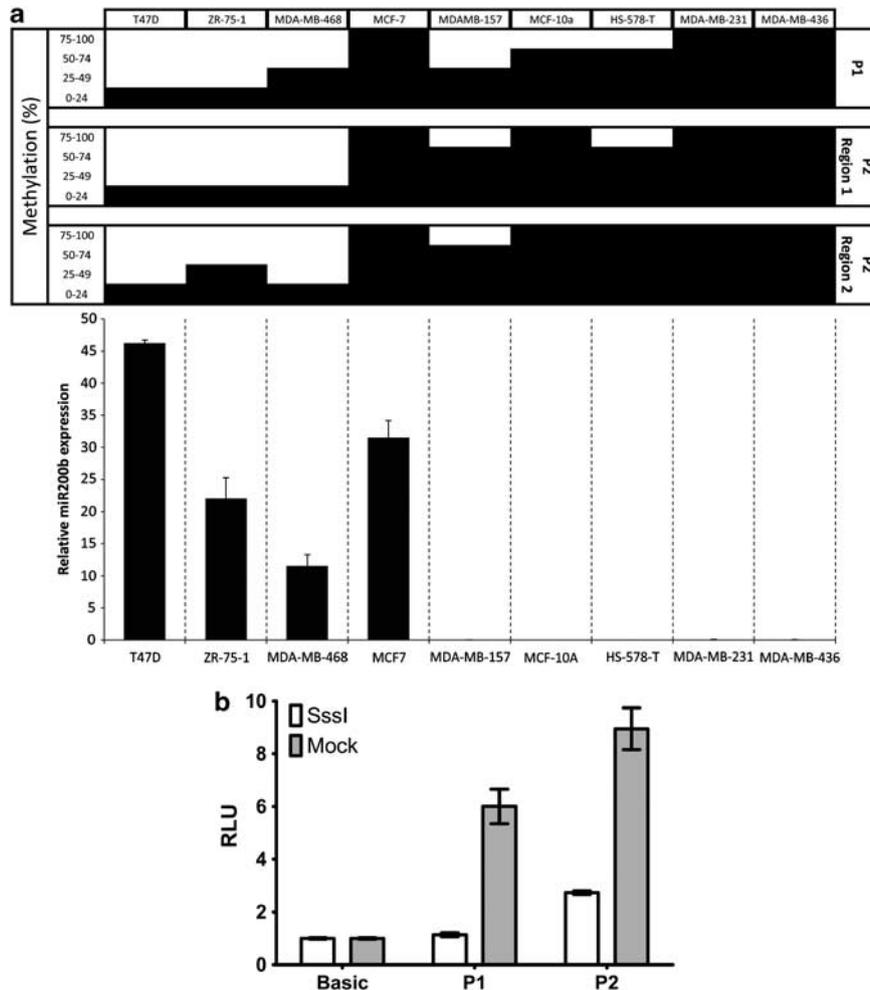


**Figure 6** miR-200b cluster has two functional promoters. (a) Promoter activities of miR-200b P1 and P2 in MCF7 and MDA-MB-231 cells. Reporter activity is expressed in RLU  $\pm$  the s.e.m. Data were generated from three independent experiments. White bars represent activity in MDA-MB-231 cells whereas grey bars represent activity in MCF7 cells. (b) Luciferase activities of various miR-200b P2 5' and 3' truncations. Graphical representation of the various reporter fragments in relation to their genomic locations and their associated reporter activities expressed in RLU. Error bars represent s.e. of three separate experiments. (c) Top: relative location of the minigene to P2, Hsa-mir-200b and the CGI. Grey box represents the region deleted in the minigene promoter KO construct. Bottom: MDA-MB-231 cells were transfected with pGL3-basic, miR200b minigene or miR200b minigene promoter KO plasmids and assayed for mature miR-200b expression by TaqMan real-time PCR  $\pm$  s.e. (d) Top: schematic of the miR-200b P2 luciferase construct. The white arrow represents the P2 promoter followed by the luciferase gene. Primers R1, R2, A to D are represented small arrows. Bottom: DNA agarose gel of the 5' PCR walk using the R2 primer with either Primer A, B, C or D, as labelled above each lane. (e) RNA-sequence profiles in T47D and MCF7 and RNA Polymerase II (RNA PolII) chip profile of MCF7 at the miR-200b locus. Peaks in RNA-sequence tracks represent expression detected at that region. Peaks represent RNA PolII binding in RNA PolII track. Horizontal bars indicate the location of the miR-200b cluster, CGI and oestrogen response element.

clinical samples. In all cases, P1 and P2 were differentially methylated in both tumours and lymph nodes (Figures 8a and b). In addition, P1, but not P2, was hypermethylated in lymph nodes compared with matched primary tumours (Figures 8c and d).

To determine if hypermethylation was associated with expression of the miR-200b cluster in primary tumours, qPCR for miR-200b was performed on tumour samples

from which RNA was available. P1 hypermethylation was associated with loss of miR-200b expression in seven out of nine samples (Supplementary Figure 3A) whereas P2 was found to be associated with loss of miR-200b expression in six out of seven samples tested (Supplementary Figure 3B). These suggested that hypermethylation of miR-200b cluster promoters could regulate miRNA expression in tumours.



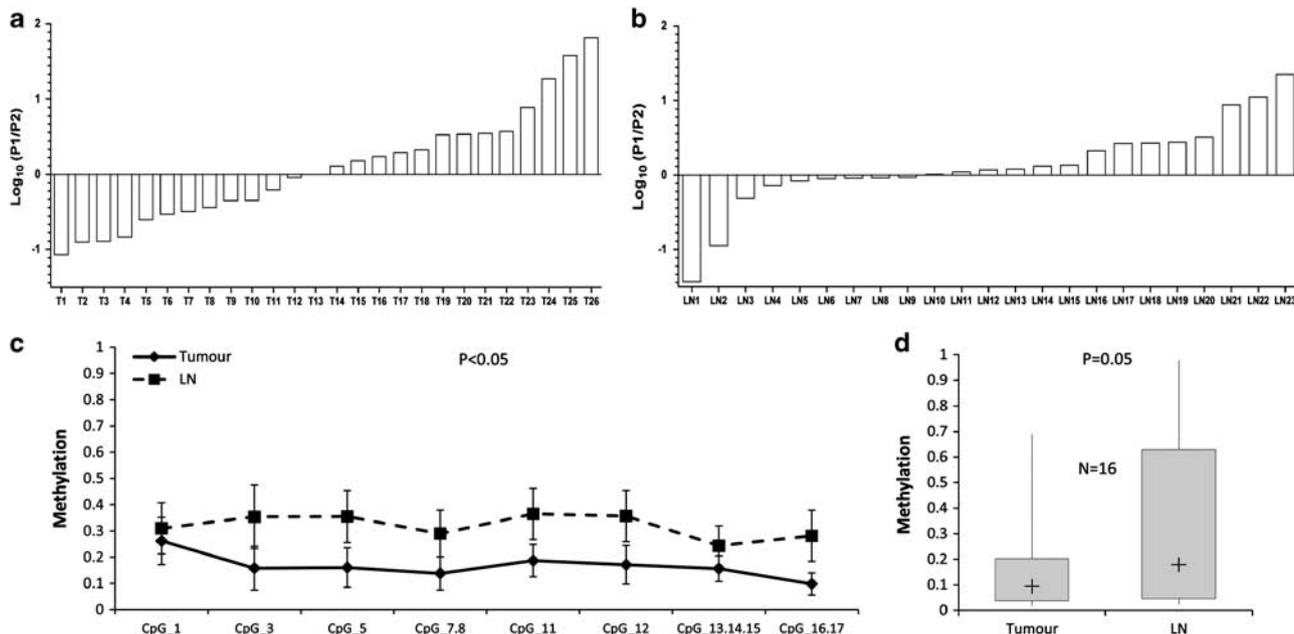
**Figure 7** DNA methylation represses miR-200b P1 and P2 activity in breast cancer cells. (a) DNA methylation and miR-200b expression levels in a panel of nine breast cancer cell lines. Top: black bars represent the percentage methylation. Bottom: miR-200b expression was assessed by qPCR. Expression is shown relative to RNU6B and bars represent the mean  $\pm$  s.e. of two independent experiments. (b) Reporter activity of miR-200b P1 and P2 methylated by *SssI* DNA methylase (white) compared with mock methylated plasmids (grey)  $\pm$  s.d. of two separate experiments.

### *Methylation of the miR-200b P2 promoter is associated with ER, PR, HER2 and androgen receptor expression in primary breast tumours*

To ascertain whether DNA methylation of the miR-200b cluster promoters is associated with expression of routinely used breast cancer biomarkers, ER, PR and HER2, methylation was assessed in patients positive and negative for expression of these receptors. Methylation of P2, but not P1, was significantly higher in tumours that were ER or PR negative (Figures 9a and b, respectively). Hypermethylation of P2 was also associated with HER2 positivity (Figure 9c). Androgen receptor, a potential breast cancer biomarker (Hu *et al.*, 2011) and regulator of the miR-200 family, (Xu *et al.*, 2010; Waltering *et al.*, 2011) was also associated with hypermethylation of P2 (Figure 9d). Although the miR-200b cluster is involved in metastasis, which in turn affects prognosis, no evidence of an association between DNA methylation and survival was found.

### Discussion

Transcriptional regulation of miRNA genes is poorly understood and only a few miRNA promoters have been reported. A comprehensive understanding of miRNA promoters is a prerequisite for their use as genetic or epigenetic biomarkers. In this report, novel CGI-associated miRNA promoters were mapped and analysed for associations between DNA methylation and miRNA expression. In all, 59% of the miRNAs examined were associated with a CGI within 5kb upstream, similar to the estimated proportion of CGI-associated coding genes and was consistent with previous estimations for miRNA promoters (Ozsolak *et al.*, 2008; Corcoran *et al.*, 2009). Twenty-two novel promoters were identified and shown to be active in reporter assays. MiR-10b had a previously described promoter (Ma *et al.*, 2007; Zhou *et al.*, 2007) immediately upstream of the mature miRNA sequence (Figure 2). However, we could



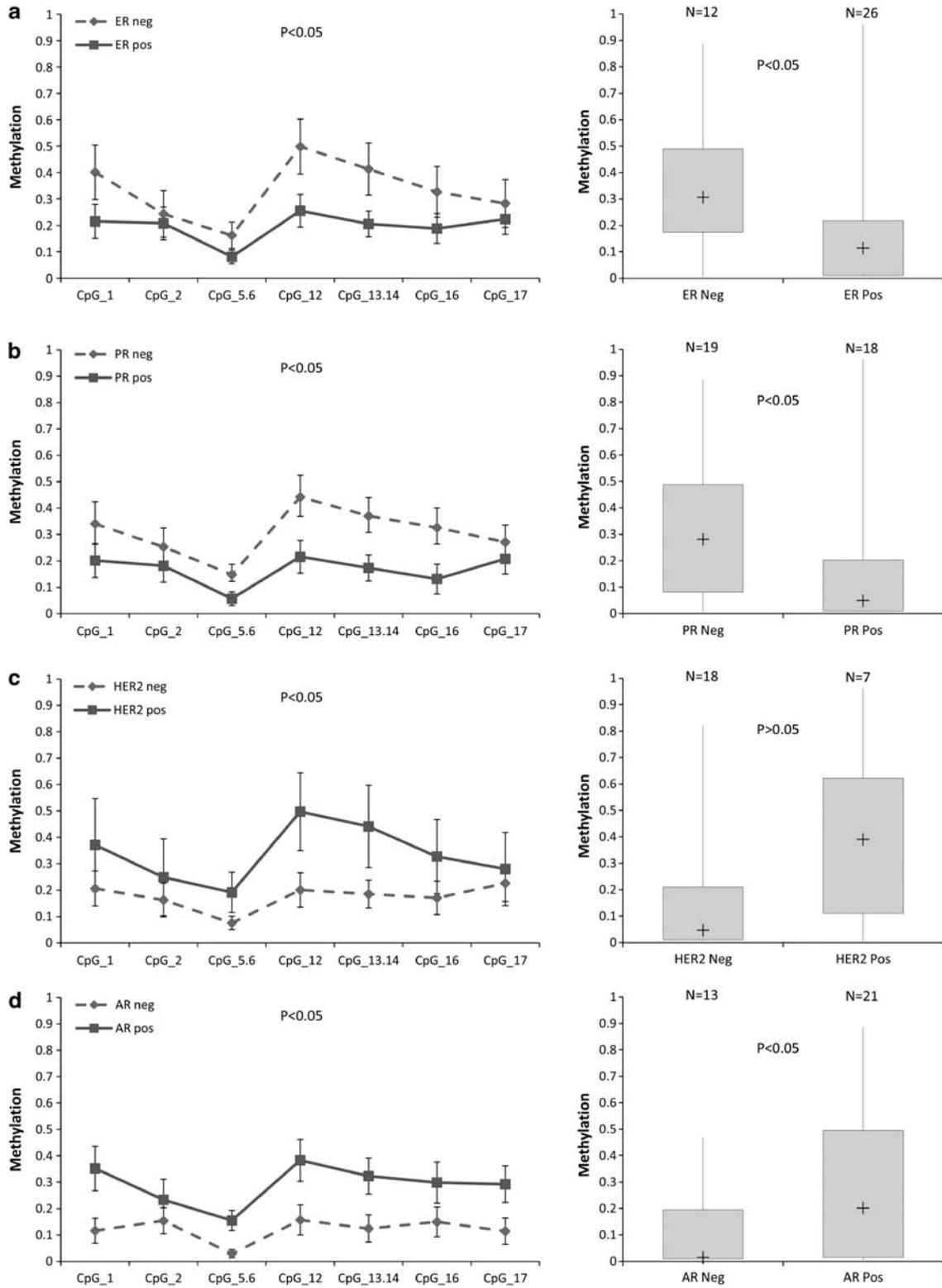
**Figure 8** Differential methylation of P1 and P2 in clinical samples. (a) Log<sub>10</sub> ratios of P1 to P2 in 26 primary tumours. (b) Log<sub>10</sub> ratios of P1 to P2 in 23 lymph nodes (LN). Positive values: P1 > P2; negative values: P1 < P2. Graph heights represent magnitude of difference in methylation between P1 and P2. (c) Mean methylation profile in matched tumours and LN with horizontal s.e. bars for individual CpG units. *t*-Test *P*-value as indicated. (d) Box plot of the average methylation in tumours and LN. (+): median, box: 25–75 percentile, whiskers: max/min, N: sample size, Mann–Whitney *P*-values as indicated.

not detect promoter activity for this fragment in the breast cancer cells tested (Figure 4c). We were also unable to detect any activity in fragments encompassing CoreBoost<sub>HM</sub> predicted promoter regions for miR-125a in the cell lines used (Figures 3 and 4i). A likely explanation was that neither cell line expressed miR-125a.

In all, 7 of 15 miRNAs had two or more promoters in close proximity, usually at either ends of the associated CGI. Although it was not clear how the multiple promoters function in regulating their miRNA genes or why the promoters were usually at either end of the CGI, it was evident that regulation of miRNAs is a complex process. The miR-200b cluster (miR-200a, miR-200b and miR-429) is an example of a miRNA with promoters at either end of the CGI. The P1 promoter, located at the distal end of the CGI, was predicted (Bracken *et al.*, 2008) based on the presence of a 5' EST, the presence of E-Box motifs and the presence of a CGI, which is commonly associated with promoters of coding genes. Further, a 7.5-kb primary transcript of the miR-200b cluster was described using a 'PCR walk' approach and P1 promoter activity was demonstrated using a luciferase reporter assay. Using a similar approach, a novel promoter, P2, is described here. P2 was predicted ~2.5 Kb downstream of P1 (Figure 3) by the CoreBoost<sub>HM</sub> promoter prediction algorithm (Wang *et al.*, 2009), which utilizes empirical data such as ESTs, RNA PolII binding and histone modification profiles in addition to DNA motifs associated with core promoters to accurately predict active promoter sites. We demonstrate that the P2 promoter has an activity similar to that of the P1 promoter (Figure 6a), is functional in breast cancer cell lines (Figures 6a and 7b)

and is able to drive the expression of miR-200b in its endogenous genomic context (Figure 6c). Thus, the P2 promoter is likely to be important in the regulation of the miR-200b cluster. Deletion of the P2 minimal promoter also reduced miR-200b levels by 50% (Figure 6c), and may indicate multiple TSS as previously suggested (Wiklund *et al.*, 2011). In addition, DNA methylation of both miR-200b promoters repressed miR-200b expression in eight out of nine breast cancer cell lines studied (Figure 7), suggesting regulation by DNA methylation. However, the precise role of P1 and P2 in regulating the cluster is unclear. In our reporter assays, P1 and P2 seemed to function independently (Figure 6a). In clinical samples, DNA methylation at P1 was also different compared with P2 in both tumour and lymph node metastases (Figures 8a and b), thus supporting the hypothesis that the two promoters have different regulatory roles. This hypothesis is supported by other studies in bladder cancer cells, where a region encompassing P2, but not P1, was unmethylated and expressed high levels of miR-200b (Wiklund *et al.*, 2011). Taken together, the evidence suggests that the P1 and P2 transcripts are regulated by different mechanisms and this could in turn have a role in regulating metastasis.

Of the eight cell lines studied, MCF7 did not show an inverse association between methylation and miRNA expression (Figure 7a). A minority of patients also did display a reciprocal relationship between promoter methylation and miRNA levels (Supplementary Figure 3). In previous studies (Han *et al.*, 2007; Wiklund *et al.*, 2011), miRNA repression by DNA methylation is usually accompanied by histone modifications associated with gene silencing. Thus, other mechanisms



**Figure 9** P2 methylation is associated with ER, PR, HER2 and AR receptor status. (a) Methylation status of ER positive (pos) and ER negative (neg) cohorts. (b) Methylation status of PR pos and PR neg cohorts. (c) Methylation status of HER2 pos and HER2 neg cohorts. (d) Methylation status of AR pos and AR neg cohorts. Left: mean methylation profile with horizontal s.e. bars for individual CpG units. *t*-Test *P*-value as indicated. Right: box plot of the average methylation in tumours and LN. (+): median, box: 25–75 percentile, whiskers: max/min, N: sample size, Mann–Whitely *P*-values as indicated.

including chromatin remodeling or post-transcriptional regulatory events may account for this inconsistency. Perhaps these repressive histone marks were absent in these cases thus resulting in open chromatin that was readily expressed.

In this study we describe, for the first time, differential methylation of the P1 and P2 region of the miR-200b cluster in breast cancer. The differential methylation is functional is evidenced by our observations that DNA methylation is inversely associated with miR-200b expression in both breast cancer cell lines (Figure 7) and clinical samples (Supplementary Figure 3). These are consistent with the previously reported tumour suppressive role of miR-200b (Korpál and Kang, 2008). They are also consistent with previous reports of aberrant DNA methylation of the miR-200b cluster proximal CGI, containing both P1 and P2, in colon, bladder and pancreatic cancers (Han *et al.*, 2007; Li *et al.*, 2010; Wiklund *et al.*, 2011).

Loss of ER and PR expression was also associated with DNA methylation at P2 in breast tumours (Figures 9a and b). Patients with tumours that express these receptors often have a better prognosis because they respond well to treatments such as Tamoxifen. We hypothesize that methylation at P2, is likely to be associated with a lower level of miRNA expression, resulting in a more aggressive tumour (Korpál and Kang, 2008) that is unresponsive to these therapies and generally associated with poor prognosis. Using publicly available ER Chip-sequence data (Schmidt *et al.*, 2010), ER bound to a putative ER response element just downstream of P1 upon ER stimulation in MCF7s (Figure 6e). In a microarray study (Klinge, 2009), miR-200a and miR-200b were significantly upregulated in MCF7 after 6 h of E<sub>2</sub> induction. However, in a similar independent study, miR-200a and 200c were found to be significantly downregulated after 48 h of E<sub>2</sub> induction (Maillot *et al.*, 2009). Although the studies seemed to have conflicting conclusions, they do suggest a possible regulatory mechanism between ER and the miR-200 family.

A double negative feedback regulatory relationship between the miR-200 family and ZEB1 (Bracken *et al.*, 2008; Burk *et al.*, 2008) has been shown to regulate the delicate balance between mesenchymal and epithelial cellular states. Based on this data, we propose that miR-200b is repressed in the early stages of tumourigenesis in order to promoter EMT and thus the spread of the tumour, followed by later induction of miR-200b to promote mesenchymal–epithelial transition and thus establishment of the tumour cells at a distant site (for example, lymph node). Our data is consistent with this as we show only P1 was hypermethylated in matched lymph nodes compared with their primary tumours. This coupled with miRNA repression, suggests a DNA methylation mechanism for EMT initiation in addition to the previously described TGFB/ZEB pathway. At P2, no differential methylation between primary tumours and matched lymph node and thus possibly maintaining base levels of miR-200 is consistent with the mesenchymal–epithelial transition observed in mouse models. Metastatic murine breast cancer cells expressing low

levels of miR-200 were able to invade distant tissue but unable to colonize. However, when miR-200 was over-expressed, these cells could form macroscopic tumours at distant sites (Dykxhoorn *et al.*, 2009). Further support for this model comes from studies in the bladder cancer (Wiklund *et al.*, 2011) where hypomethylation of the P2 region was sufficient for miR-200b cluster expression. This hypomethylation could also possibly account for the elevated levels of the miR-200 family, observed in other cancers (Hiroki *et al.*, 2010; Li *et al.*, 2010; Lee *et al.*, 2011).

Collectively, the evidence presented here indicates that miR-200b cluster regulation is complex and is regulated transcriptionally by at least two distinct promoters that are sensitive to DNA methylation. The novel P2 promoter functions independently of P1 and can drive the expression of miR-200b. However, the precise roles of P1 or P2 and under what conditions they are utilized is still not clear and will require further examination.

## Materials and methods

### Bioinformatics

A list of 93 miRNAs implicated in breast cancer was generated by literature review. Genomic sequences +5 and –1 kb of the 5' stemloop of each miRNA were analysed for CGIs using CpGPlot and CpG Island Searcher. Putative promoters are defined by a CoreBoost<sub>HM</sub> score of at least 0.5 (Wang *et al.*, 2009) within this 6 kb window. Initial 600–1000 bp fragments overlapping the predicted sites were cloned and assayed for promoter activity as described later. This process is illustrated in Figure 1.

RNA Polymerase II Chip-sequence data mapped to human genome HG18 was obtained from the National Center for Biotechnology Gene Expression Omnibus, GEO accession number GSE14664. RNA-sequence data (Wang *et al.*, 2008) mapped to human genome HG18. RNA Polymerase II and RNA-sequence data were visualized on integrative genome viewer using the data ranges indicated.

### Cell culture

Breast cancer cell lines MDAMB157, MDAMB231, MDAMB436, MDAMB468, MCF7, T47D, ZR75-1, Hs578T and BT549 were obtained from American Type Culture Collection (ATCC, Manassas, VA, USA) and cultured according to the manufacturer's recommendations.

### Patient samples

Human breast tumours and matching lymph node metastases were collected from 56 patients, as approved by local Human Ethics committees, who underwent surgical resection and did not undergo preoperative radiochemotherapy at Princess Alexandra Hospital between 1988 and 2000. All patients were female aged from 30 to 94 years old, with a median age of 56 years. ER, PR and HER2 receptor status of each patient were determined by a qualified pathologist. Details are provided in the Supplementary Information.

### DNA extractions and purifications

Genomic DNA from cell lines was extracted using the NucleoSpin Tissue Prep kit (Macherey-Nagel, Germany) according to the manufacturer's instructions. Plasmid DNA

was purified using the Miniprep Kit (Qiagen, Doncaster, VIC, Australia). For human tumour samples, four FFPE tumour-rich tissue cores (1 × 0.6 mm) were crushed and digested with proteinase K at 55 °C for 2 days. Genomic DNA was purified using the PureGene kit (Qiagen) according to the manufacturer's instructions.

#### *Generation of plasmid constructs*

All promoter reporter constructs were cloned into pGL3-Basic (Promega, Sydney, NSW, Australia) unless otherwise specified. For the *in vitro* methylation plasmids, P1 and P2 fragments were cloned into a CpG-free luciferase reporter construct pCpG-basic (Klug and Rehli, 2006; a gift from Klug and Rehli). PCR was performed using KapaHiFi polymerase (Kapa Biosystems, Woburn, MA, USA). All constructs were confirmed by sequencing. All primers used and cloning details are provided in the Supplementary.

#### *Transfections and reporter assays*

All transfections used a 3-μl:1 μg ratio of Fugene (Roche, Castle Hill, NSW, Australia) transfection reagent to DNA. For luciferase assays, either MCF7 or MDA-MB-231 cells were co-transfected with 400 ng of promoter construct and 10 ng of RL-TK plasmid (Promega) as a transfection control and harvested and assayed for reporter activity after 48 h. The Dual-Glo luciferase Assay kit (Promega) was used as recommended by the manufacturer. Firefly luciferase levels were normalized to *Renilla* luciferase levels and expressed relative to pGL3-basic levels (RLU). Statistical analysis was performed using unpaired two-tailed *t*-test.

For minigene experiments, MDA-MB-231 cells were grown to 60–70% confluence in 6-well plates, and transfected with 1 μg of DNA and harvested after 72 h.

#### *Identification of TSS of Hsa-mir-200b*

Total RNA from MCF7 transfected with the P2 luciferase reporter construct was extracted using Trizol (Invitrogen) and DNaseI (NEB, Ipswich, MA, USA) treated. First strand cDNA was reverse transcribed using SuperScriptIII (Invitrogen) using a luciferase specific primer, R1, at 50 °C. This then served as a template for PCR amplification. PCR 'walking' towards the 5' end was performed using primers A to D with R2. All PCR products were visualized on a 1% agarose gel. Primer sequences are given in Supplementary Table 3. Classical 5'RACE was performed as previously described (2005).

#### *Quantitation of miRNAs*

Total RNA was extracted from cell lines using Trizol (Invitrogen). RNA from clinical samples was extracted using miRNeasy kit (Qiagen). For miR200b and miR335 experiments, cDNA was made from total RNA using TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems, Mulgrave, VIC, Australia) with both reverse transcription miRNA and RNU6B (loading control) primers in the same reaction. Real-time PCR was performed using the TaqMan microRNA Assay (Applied Biosystems) according to the manufacturer's instructions. For all other miRNAs, the Qiagen miScript PCR system for miRNA quantification was used with the RNU6B loading control. Changes in expression levels were calculated using  $\Delta\Delta C_t$  method (Livak and Schmittgen, 2001).

#### *Bisulfite modification and methylation-sensitive high-resolution melt analysis*

2 μg of DNA extracted from cell lines was subjected to bisulfite modification with MethylEasy Xceed kit (Human Genetic Signatures, Randwick, NSW, Australia) according to manu-

facturer's instructions. PCR amplification and methylation-sensitive high-resolution melt analysis (Wojdacz and Dobrovic, 2007) was performed in duplicate on the RotorGene Q (Qiagen). Primers were designed according to the principles outlined (Wojdacz and Dobrovic, 2007) to control for PCR bias and are shown in Supplementary Tables 4 and 5. PCR conditions are provided in the Supplementary. Bisulfite treated CpGenome Universal Methylated DNA (Chemicon, Millipore, Kilsyth, VIC, Australia) and DNA from the appropriate cell lines were used as positive/methylated and negative/unmethylated controls, respectively. WGA DNA made with the GenomiPhi kit (Amersham GE Healthcare, Rydalmere, NSW, Australia) was used as unmethylated controls for miR335 and miR663. Included in the analysis of each region, controls were mixed in 25, 50 and 75% methylated to unmethylated template ratios.

#### *In vitro methylation of plasmid DNA*

DNA was methylated using *SssI* (NEB) as previously described (Klug and Rehli, 2006). Briefly, plasmids were incubated with *SssI* (2.5 U/μg) with 160 μM S-adenosylmethionine at 37 °C for 4 h and supplemented with an additional 160 μM of S-adenosylmethionine for another 4 h at 37 °C. Mock methylated plasmids controls were treated similarly but without enzyme. Plasmids were recovered by phenol/chloroform, followed by ethanol precipitation, transfected into T47D cells and luciferase assays performed.

#### *Sequenom MassArray*

Genomic DNA from clinical samples was bisulfite converted with EZ-96 DNA methylation kit (Zymo Research, Irvine, CA, USA). Methylation levels in clinical samples were determined using Sequenom MassArray, performed according to manufacturer's recommendations for T-cleavage chemistry protocol and analysed between a 1640 and 7000 mass window (Coolen *et al.*, 2007). Average methylation of each patient is defined as the average percent methylation of all CpG units in each amplicon. Average methylation of each CpG cluster (or profile) is defined as the average percent methylation of the cohort for that specific CpG cluster. In all, 0–100% methylation are represented by 0.0 to 1.0. Primer sequences are provided in Supplementary Table 4.

#### **Conflict of interest**

The authors declare no conflict of interest.

#### **Acknowledgements**

This project was funded by grants from the National Breast Cancer Foundation (NBCF), Cancer Australia and National Health and Medical Research Council (NHMRC). We are grateful to Dr Maja Klug for generously providing the CpG-Less reporter plasmid.

#### *Web Resources*

CpG plot: <http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html>

CpG Island Searcher: <http://cpgislands.usc.edu/>

CoreBoost\_HM: [http://rulai.cshl.edu/tools/CoreBoost\\_HM/RNA-seq\\_data](http://rulai.cshl.edu/tools/CoreBoost_HM/RNA-seq_data): <http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/>.

Integrative Genome Viewer: <http://www.broadinstitute.org/igv>

## References

- (2005). Rapid amplification of 5' complementary DNA ends (5' RACE). *Nat Methods* **2**: 629–630.
- Bartel DP. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–233.
- Bartels CL, Tsongalis GJ. (2009). MicroRNAs: novel biomarkers for human cancer. *Clin Chem* **55**: 623–631.
- Boyd SD. (2008). Everything you wanted to know about small RNA but were afraid to ask. *Lab Invest* **88**: 569–578.
- Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, Shannon MF *et al.* (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res* **68**: 7846–7854.
- Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S *et al.* (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* **9**: 582–589.
- Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. (2007). Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res* **35**: e119.
- Corcoran DL, Pandit KV, Gordon B, Bhattacharjee A, Kaminski N, Benos PV. (2009). Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. *PLoS One* **4**: e5279.
- Cullen BR. (2004). Transcription and processing of human microRNA precursors. *Mol Cell* **16**: 861–865.
- Dykxhoorn DM, Wu Y, Xie H, Yu F, Lal A, Petrocca F *et al.* (2009). miR-200 enhances mouse breast cancer cell colonization to form distant metastases. *PLoS One* **4**: e7181.
- Foekens JA, Sieuwerts AM, Smid M, Look MP, de Weerd V, Boersma AW *et al.* (2008). Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. *Proc Natl Acad Sci U S A* **105**: 13021–13026.
- Fujita S, Iba H. (2008). Putative promoter regions of miRNA genes involved in evolutionarily conserved regulatory systems among vertebrates. *Bioinformatics* **24**: 303–308.
- Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G *et al.* (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**: 593–601.
- Han L, Witmer PD, Casey E, Valle D, Sukumar S. (2007). DNA methylation regulates MicroRNA expression. *Cancer Biol Ther* **6**: 1284–1288.
- Hiroki E, Akahira J, Suzuki F, Nagase S, Ito K, Suzuki T *et al.* (2010). Changes in microRNA expression levels correlate with clinicopathological features and prognoses in endometrial serous adenocarcinomas. *Cancer Sci* **101**: 241–249.
- Hu R, Dawood S, Holmes MD, Collins LC, Schnitt SJ, Cole K *et al.* (2011). Androgen receptor expression and breast cancer survival in postmenopausal women. *Clin Cancer Res* **17**: 1867–1874.
- Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S *et al.* (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* **65**: 7065–7070.
- Klinge CM. (2009). Estrogen regulation of microRNA expression. *Curr Genomics* **10**: 169–183.
- Klug M, Rehli M. (2006). Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. *Epigenetics* **1**: 127–130.
- Korpala M, Kang Y. (2008). The emerging role of miR-200 family of microRNAs in epithelial-mesenchymal transition and cancer metastasis. *RNA Biol* **5**: 115–119.
- Lee JW, Park YA, Choi JJ, Lee YY, Kim CJ, Choi C *et al.* (2011). The expression of the miRNA-200 family in endometrial endometrioid carcinoma. *Gynecol Oncol* **120**: 56–62.
- Lehmann U, Hasemeier B, Christgen M, Muller M, Romermann D, Langer F *et al.* (2008). Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J Pathol* **214**: 17–24.
- Li A, Omura N, Hong SM, Vincent A, Walter K, Griffith M *et al.* (2010). Pancreatic cancers epigenetically silence SIP1 and hypomethylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels. *Cancer Res* **70**: 5226–5237.
- Linhart C, Halperin Y, Shamir R. (2008). Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. *Genome Res* **18**: 1180–1189.
- Livak KJ, Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402–408.
- Lodygin D, Tarasov V, Epanchintsev A, Berking C, Knyazeva T, Korner H *et al.* (2008). Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle* **7**: 2591–2600.
- Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C *et al.* (2009). MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Res* **11**: R27.
- Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F *et al.* (2007). Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res* **67**: 1424–1429.
- Ma L, Teruya-Feldstein J, Weinberg RA. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449**: 682–688.
- Maillot G, Lacroix-Triki M, Pierredon S, Gratadou L, Schmidt S, Benes V *et al.* (2009). Widespread estrogen-dependent repression of microRNAs involved in breast tumor cell growth. *Cancer Res* **69**: 8332–8340.
- Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S *et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**: 521–533.
- Mattie MD, Benz CC, Bowers J, Sensinger K, Wong L, Scott GK *et al.* (2006). Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Mol Cancer* **5**: 24.
- Mirnezami AH, Pickard K, Zhang L, Primrose JN, Packham G. (2009). MicroRNAs: key players in carcinogenesis and novel therapeutic targets. *Eur J Surg Oncol* **35**: 339–347.
- Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox KA, Xing Y *et al.* (2010). Structure and activity of putative intronic miRNA promoters. *Rna* **16**: 495–505.
- Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG *et al.* (2008). Chromatin structure analyses identify miRNA promoters. *Genes Dev* **22**: 3172–3183.
- Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS *et al.* (2010). A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* **20**: 578–588.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**: 10869–10874.
- Visone R, Croce CM. (2009). MiRNAs and cancer. *Am J Pathol* **174**: 1131–1138.
- Waltering KK, Porkka KP, Jalava SE, Urbanucci A, Kohonen PJ, Latonen LM *et al.* (2011). Androgen regulation of microRNAs in prostate cancer. *Prostate* **71**: 604–614.
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C *et al.* (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.
- Wang X, Xuan Z, Zhao X, Li Y, Zhang MQ. (2009). High-resolution human core-promoter prediction with CoreBoost\_HM. *Genome Res* **19**: 266–275.
- Wiklund ED, Bramsen JB, Hulf T, Dyrskjot L, Ramanathan R, Hansen TB *et al.* (2011). Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. *Int J Cancer* **128**: 1327–1334.

- Wojdacz TK, Dobrovic A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res* **35**: e41.
- Xu G, Wu J, Zhou L, Chen B, Sun Z, Zhao F *et al.* (2010). Characterization of the small RNA transcriptomes of androgen dependent and independent prostate cancer cell line by deep sequencing. *PLoS One* **5**: e15519.
- Yan HL, Xue G, Mei Q, Wang YZ, Ding FX, Liu MF *et al.* (2009). Repression of the miR-17-92 cluster by p53 has an important function in hypoxia-induced apoptosis. *EMBO J* **28**: 2719–2732.
- Yan LX, Huang XF, Shao Q, Huang MY, Deng L, Wu QL *et al.* (2008). MicroRNA miR-21 overexpression in human

- breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA* **14**: 2348–2360.
- Zhou X, Ruan J, Wang G, Zhang W. (2007). Characterization and identification of microRNA core promoters in four model species. *PLoS Comput Biol* **3**: e37.



**This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>**

Supplementary Information accompanies the paper on the Oncogene website (<http://www.nature.com/onc>)